W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Oct 31, 2023 – 01:46 PM JST

PDB ID	:	5EWX
Title	:	Fusion protein of T4 lysozyme and B4 domain of protein A from staphylococcal aureus with chemical cross-linker EY-CBS
Authors	:	Jeong, W.H.; Lee, H.; Song, D.H.; Lee, J.O.
Deposited on	:	2015-11-22
Resolution	:	2.60 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

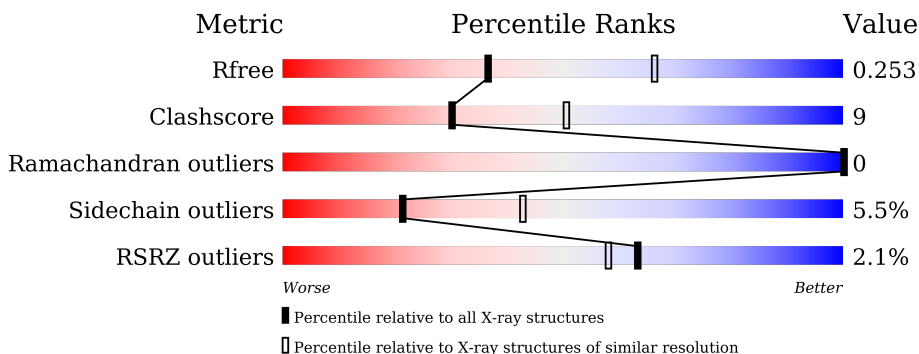|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Percentile Ranks	Value
Rfree		0.253
Clashscore		9
Ramachandran outliers		0
Sidechain outliers		5.5%
RSRZ outliers		2.1%

Worse   *Better*

▌Percentile relative to all X-ray structures

▯Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	227	4% — 79% / 11% · 7%
1	B	227	75% / 17% · 6%

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 3472 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Endolysin,Immunoglobulin G-binding protein A,Endolysin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	210	Total 1666	C 1043	N 301	O 315	S 7	0	0	0
1	B	214	Total 1682	C 1051	N 305	O 319	S 7	0	0	0

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	12	GLY	ARG	see sequence details	UNP P00720
A	1209	GLY	-	linker	UNP P00720
A	1210	GLY	-	linker	UNP P00720
A	1211	GLY	-	linker	UNP P00720
A	1212	GLY	-	linker	UNP P00720
A	1213	SER	-	linker	UNP P00720
A	1214	GLY	-	linker	UNP P00720
A	1215	GLY	-	linker	UNP P00720
A	1216	GLY	-	linker	UNP P00720
A	1217	GLY	-	linker	UNP P00720
A	1218	SER	-	linker	UNP P00720
A	1218A	VAL	ALA	engineered mutation	UNP P38507
A	1240	ALA	GLY	engineered mutation	UNP P38507
A	1258	CYS	GLU	engineered mutation	UNP P38507
A	1261	ALA	LYS	engineered mutation	UNP P38507
A	1262	ALA	LEU	engineered mutation	UNP P38507
A	40	CYS	ASN	engineered mutation	UNP P00720
A	54	THR	CYS	engineered mutation	UNP P00720
A	97	ALA	CYS	engineered mutation	UNP P00720
A	137	ARG	ILE	see sequence details	UNP P00720
A	162	ALA	LYS	engineered mutation	UNP P00720
B	12	GLY	ARG	see sequence details	UNP P00720
B	1209	GLY	-	linker	UNP P00720
B	1210	GLY	-	linker	UNP P00720
B	1211	GLY	-	linker	UNP P00720

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1212	GLY	-	linker	UNP P00720
B	1213	SER	-	linker	UNP P00720
B	1214	GLY	-	linker	UNP P00720
B	1215	GLY	-	linker	UNP P00720
B	1216	GLY	-	linker	UNP P00720
B	1217	GLY	-	linker	UNP P00720
B	1218	SER	-	linker	UNP P00720
B	1218A	VAL	ALA	engineered mutation	UNP P38507
B	1240	ALA	GLY	engineered mutation	UNP P38507
B	1258	CYS	GLU	engineered mutation	UNP P38507
B	1261	ALA	LYS	engineered mutation	UNP P38507
B	1262	ALA	LEU	engineered mutation	UNP P38507
B	40	CYS	ASN	engineered mutation	UNP P00720
B	54	THR	CYS	engineered mutation	UNP P00720
B	97	ALA	CYS	engineered mutation	UNP P00720
B	137	ARG	ILE	see sequence details	UNP P00720
B	162	ALA	LYS	engineered mutation	UNP P00720

- Molecule 2 is 2,2'-ethyne-1,2-diylbis{5-[(chloroacetyl)amino]benzenesulfonic acid} (three-letter code: EYC) (formula: $C_{18}H_{14}Cl_2N_2O_8S_2$).
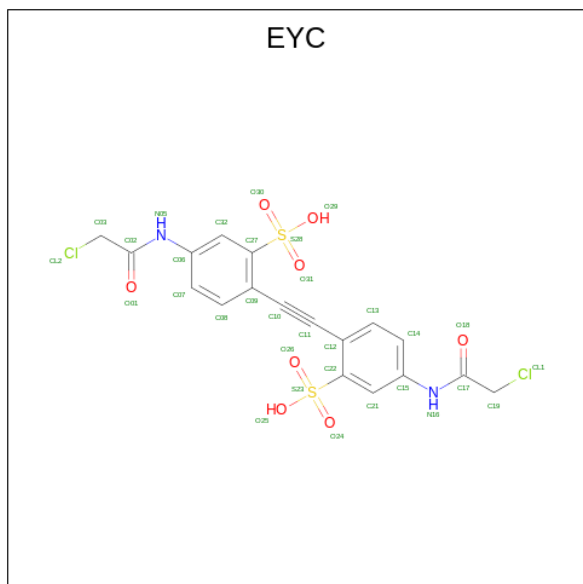
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	S	0	0
			30	18	2	8	2		
2	B	1	Total	C	N	O	S	0	0
			30	18	2	8	2		

• Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	29	Total 29	O 29	0	0
3	B	35	Total 35	O 35	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.
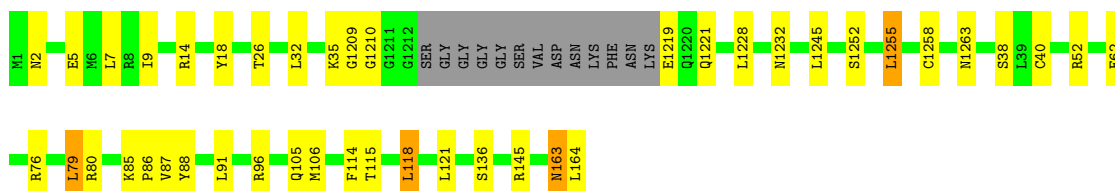
• Molecule 1: Endolysin,Immunoglobulin G-binding protein A,Endolysin
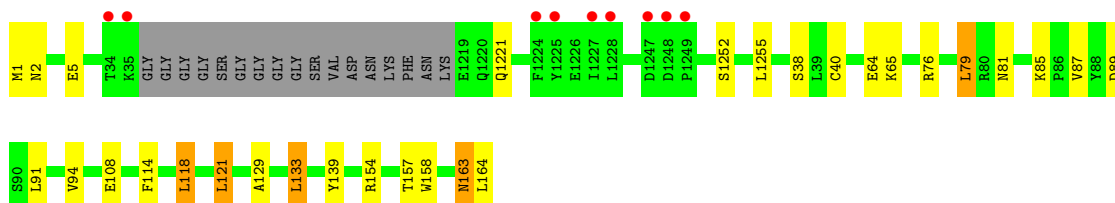
• Molecule 1: Endolysin,Immunoglobulin G-binding protein A,Endolysin

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	89.52Å 149.85Å 55.84Å 90.00° 122.69° 90.00°	Depositor
Resolution (Å)	42.74 – 2.60 42.74 – 2.59	Depositor EDS
% Data completeness (in resolution range)	94.1 (42.74-2.60) 94.8 (42.74-2.59)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.06 (at 2.58Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.203 , 0.253 0.206 , 0.253	Depositor DCC
R_{free} test set	1878 reflections (9.99%)	wwPDB-VP
Wilson B-factor (Å^2)	40.1	Xtriage
Anisotropy	0.657	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 41.0	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	3472	wwPDB-VP
Average B, all atoms (Å^2)	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.03% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: EYC

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$|Z|$ >5	RMSZ	#$|Z|$ >5
1	A	0.46	0/1690	0.63	0/2279
1	B	0.50	0/1706	0.65	1/2299 (0.0%)
All	All	0.48	0/3396	0.64	1/4578 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	7	LEU	CA-CB-CG	-5.32	103.07	115.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1666	0	1669	33	0
1	B	1682	0	1681	34	0
2	A	30	0	0	4	0
2	B	30	0	0	4	0
3	A	29	0	0	15	1
3	B	35	0	0	9	1
All	All	3472	0	3350	63	1

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 9.

The worst 5 of 63 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:40:CYS:SG	2:A:1301:EYC:C19	2.15	1.34
1:A:40:CYS:SG	2:A:1301:EYC:C17	2.29	1.19
1:A:40:CYS:SG	2:A:1301:EYC:O18	2.12	1.07
1:A:76:ARG:HD3	3:A:1402:HOH:O	1.60	0.99
1:B:1221:GLN:NE2	3:B:1404:HOH:O	1.98	0.97

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:1421:HOH:O	3:B:1417:HOH:O[1_556]	1.73	0.47

5.3　Torsion angles ⓘ

5.3.1　Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	206/227 (91%)	202 (98%)	4 (2%)	0	100	100
1	B	210/227 (92%)	206 (98%)	4 (2%)	0	100	100
All	All	416/454 (92%)	408 (98%)	8 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2　Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	174/183 (95%)	165 (95%)	9 (5%)	23	46
1	B	174/183 (95%)	164 (94%)	10 (6%)	20	41
All	All	348/366 (95%)	329 (94%)	19 (6%)	21	43

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	79	LEU
1	B	121	LEU
1	B	163	ASN
1	B	118	LEU
1	A	163	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3　RNA (i)

There are no RNA molecules in this entry.

5.4　Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5　Carbohydrates (i)

There are no monosaccharides in this entry.

5.6　Ligand geometry (i)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the

expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
2	EYC	A	1301	-	27,31,33	2.70	8 (29%)	36,46,48	2.75	13 (36%)
2	EYC	B	1301	-	27,31,33	2.37	8 (29%)	36,46,48	2.05	13 (36%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	EYC	A	1301	-	-	1/25/25/29	0/2/2/2
2	EYC	B	1301	-	-	8/25/25/29	0/2/2/2

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1301	EYC	C02-N05	6.85	1.48	1.36
2	A	1301	EYC	C12-C11	5.45	1.53	1.43
2	A	1301	EYC	C09-C10	5.39	1.53	1.43
2	B	1301	EYC	C09-C10	5.29	1.53	1.43
2	B	1301	EYC	C12-C11	4.99	1.52	1.43

The worst 5 of 26 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1301	EYC	O26-S23-C22	-9.08	91.15	106.51
2	A	1301	EYC	C13-C12-C11	5.01	129.79	120.11
2	A	1301	EYC	C21-C15-N16	-4.71	104.81	120.18
2	A	1301	EYC	C03-C02-N05	4.45	121.52	114.98
2	B	1301	EYC	C21-C15-N16	-4.41	105.77	120.18

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	1301	EYC	O18-C17-N16-C15
2	B	1301	EYC	C10-C11-C12-C22
2	B	1301	EYC	C03-C02-N05-C06
2	B	1301	EYC	O01-C02-N05-C06

Continued on next page...

Continued from previous page...

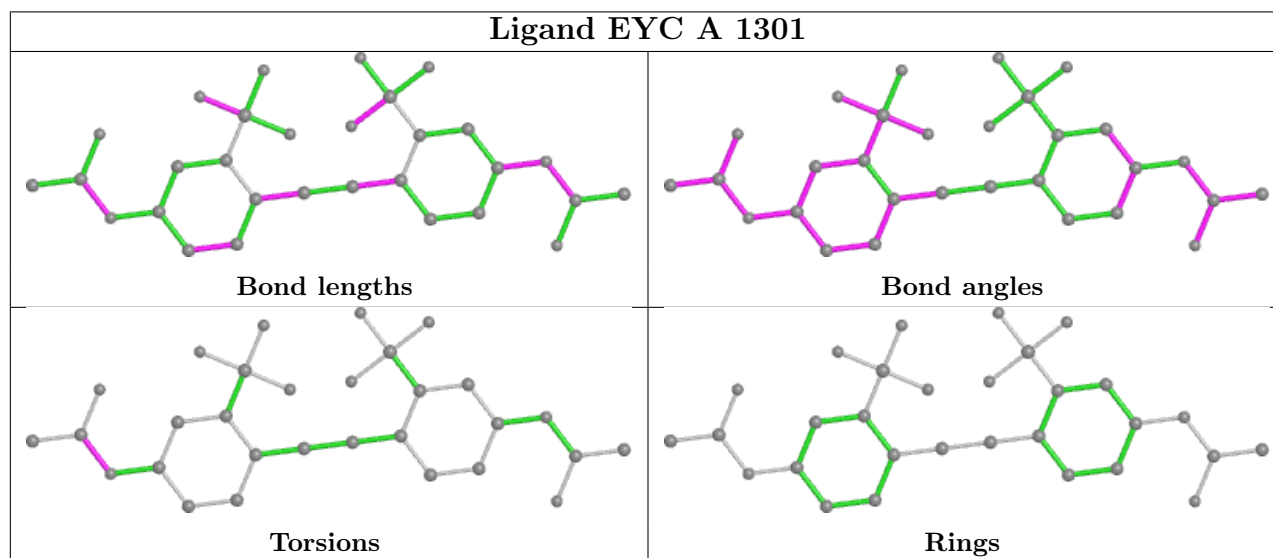| Mol	Chain	Res	Type	Atoms
2	B	1301	EYC	C10-C11-C12-C13

There are no ring outliers.

2 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1301	EYC	4	0
2	B	1301	EYC	4	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand EYC A 1301

Bond lengths

Bond angles

Torsions

Rings

Ligand EYC B 1301	
Bond lengths	Bond angles
Torsions	Rings
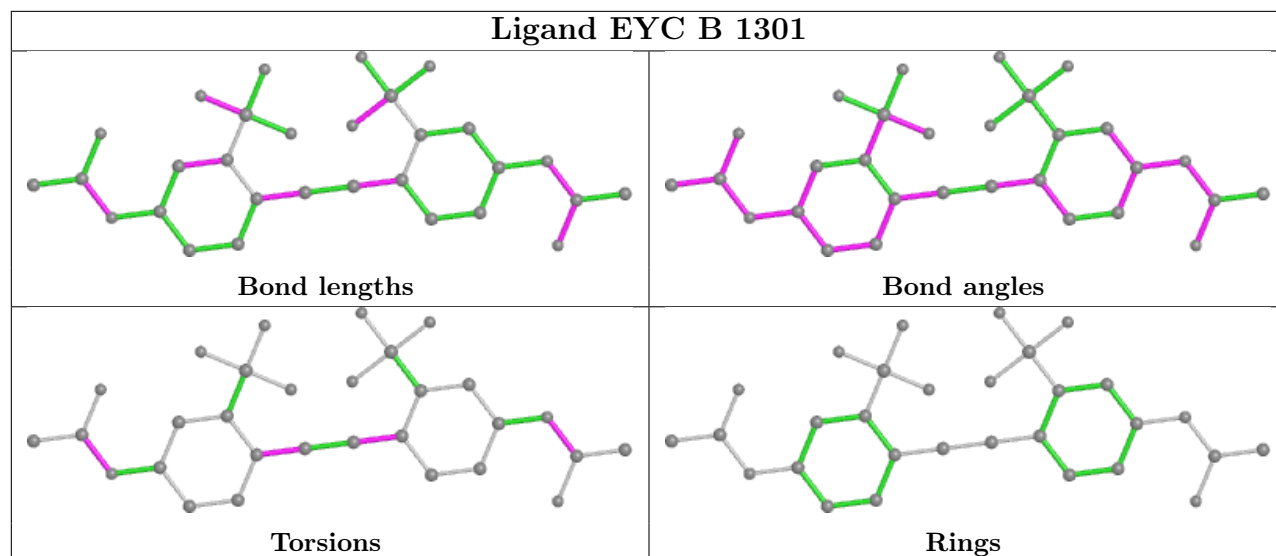
5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6   Fit of model and data ⓘ

6.1   Protein, DNA and RNA chains ⓘ

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	210/227 (92%)	0.02	9 (4%)	35	28	32, 52, 111, 130	0
1	B	214/227 (94%)	-0.17	0	100	100	31, 48, 85, 101	0
All	All	424/454 (93%)	-0.08	9 (2%)	63	58	31, 50, 97, 130	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1227	ILE	3.2
1	A	1247	ASP	3.1
1	A	1249	PRO	3.0
1	A	1225	TYR	2.9
1	A	1224	PHE	2.8

6.2   Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.
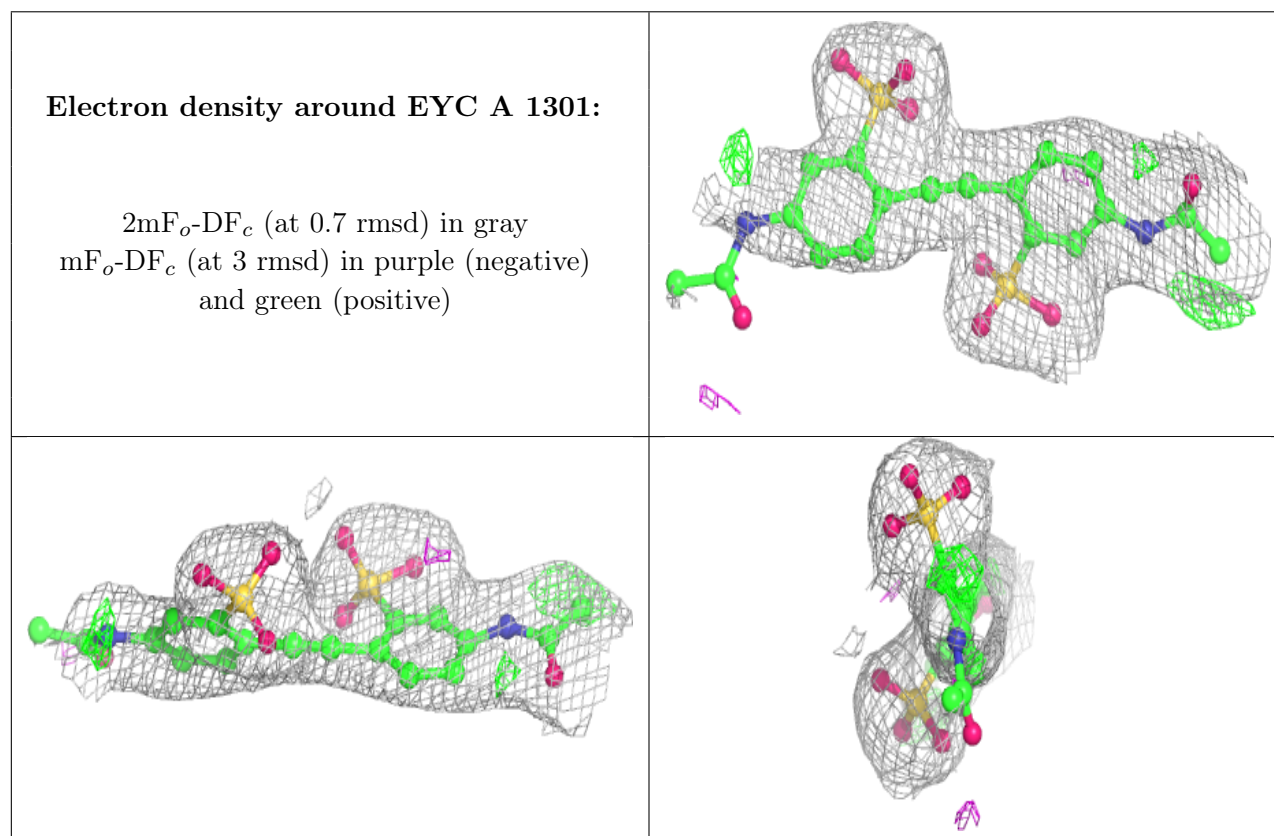
6.3   Carbohydrates ⓘ

There are no monosaccharides in this entry.

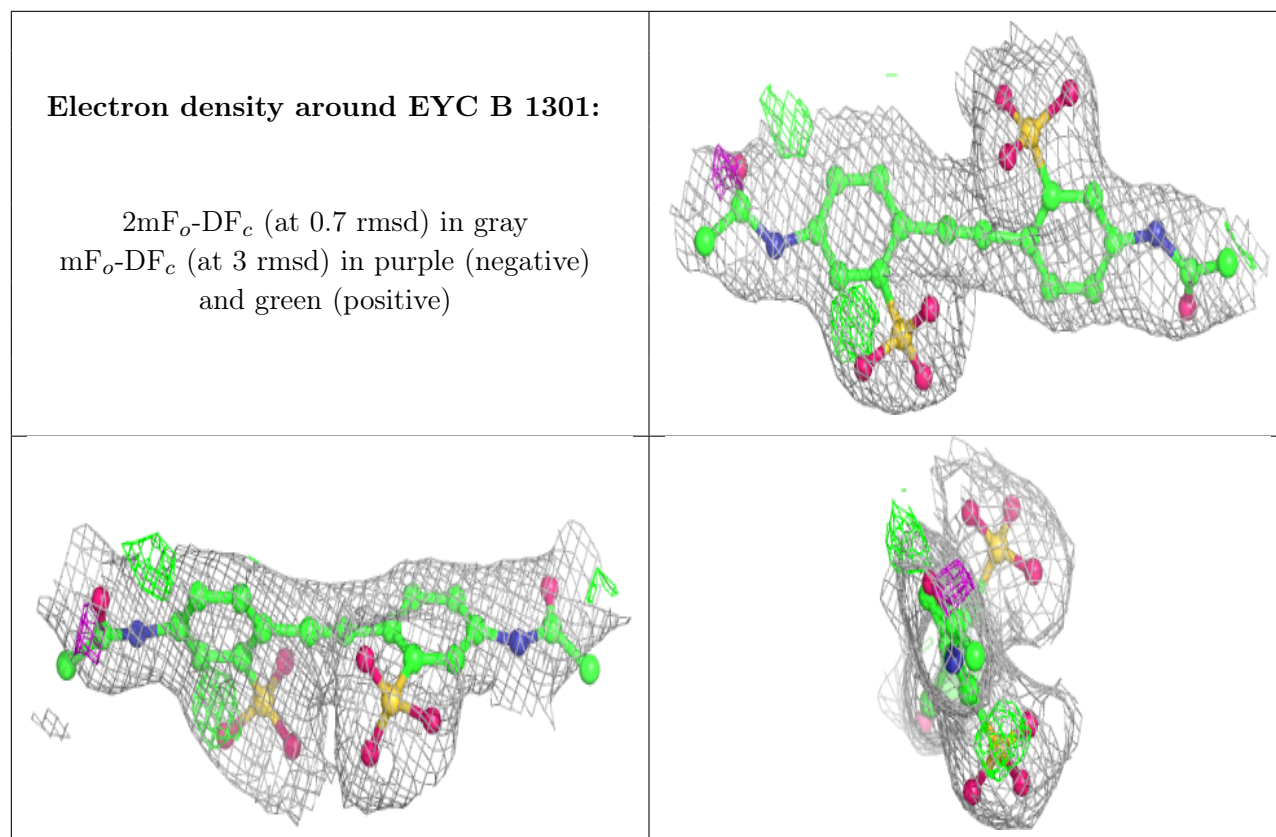## 6.4   Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	EYC	A	1301	30/32	0.94	0.21	15,58,104,111	0
2	EYC	B	1301	30/32	0.95	0.19	22,55,76,102	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around EYC A 1301:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)

Electron density around EYC B 1301:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)

6.5 Other polymers ⓘ

There are no such residues in this entry.